TERRY SANFORD INSTITUTE
OF PUBLIC POLICY
DUKE

Payments for Environmental Services: Empirical analysis for Costa Rica

Alexander Pfaff,
Juan Andres Robalino
and G. Arturo Sanchez-Azofeifa

Working Papers Series
SAN08-05

March 2008

TERRY SANFORD INSTITUTE OF PUBLIC POLICY
DUKE

Payments for Environmental Services: Empirical analysis for Costa Rica

Alexander Pfaff [1*], Juan Andres Robalino [2]
and G. Arturo Sanchez-Azofeifa [3]

Abstract

Evaluating its impact using the deforestation observed in matched untreated areas, we find that Costa Rica's 'PSA' program of payments for environmental services had little effect on 1997-2000 forest clearing. Reasons include: a low national rate of deforestation; no targeting of those locations more likely to change land usage; and a goal of transferring surplus to landowners. This pioneering effort could save much of its budget, or greatly increase forest impact from current funds, if it could avoid enrolling lands which would remain forested even without such payments.

Keywords: forest, ecosystem services, payments, Costa Rica

[1] Duke University
[2] CATIE
[3] University of Alberta

[*] Corresponding author: Alexander Pfaff
Associate Professor of Public Policy Studies
Terry Sanford Institute of Public Policy
Duke University
Durham, NC 27708-0245
alex.pfaff@duke.edu, (919) 613-9240

1. Introduction

Payments for environmental services are in principle an effective way to induce conservation while compensating those who incur its costs. Among conservation policies, they may lessen conflicts between conservation and local welfare (Ferraro 2001). Such payments appear to have global appeal (Frank and Muller 2003, Chomitz et al. 1998, Rojas and Aylward 2003, Miranda 2003, Echaverria et al. 2004, Tikka 2003, Rosales 2003, Smith 1995, and Szentandrasi 1995).

However, these and other more common conservation policies such as forest reserves or parks are rarely subject to rigorous impact evaluation. Andam et al. 2007, for instance, claims to be the first park-impact evaluation to take into account the non-random distribution of parks due to the processes that drive reserve site choice. More generally, Ferraro and Pattanayak 2006 call for empirical evaluation of conservation actions. They note that forest impact is complicated because to measure "avoided deforestation" involves a counterfactual that analysts must construct – the deforestation that would have occurred if an area of forest had not received policy protection.

In this vein, we ask whether ecopayments have significantly lowered deforestation. If not, few of their potential benefits will obtain. We empirically examine forest-protection contracts in Costa Rica's environmental services payments program (the PSA; or Pagos por Servicios Ambientales). It is one of the most advanced initiatives of its kind within the developing world (Pagiola 2002) and is widely noted and studied (Rojas and Aylward 2003). We find that it has had little impact upon deforestation, for reasons in part widely relevant but in part special to Costa Rican history.

Costa Rican deforestation decreased significantly during the 1980s and 1990s (de Camino et al. 2000, Sanchez et al. 2003). Many assert or assume that the PSA contributed significantly. However, the payments program started in 1997 and Forestry Law 7575, without compensation, greatly restricted post-1997 forest clearing while ecotourism has flourished (Rojas and Aylward 2003), ranching was hurt by falling prices, and more generally other Costa Rican conservation policies and other shifts over time appear to have lowered forest clearing (Kerr and Pfaff 2007). As the PSA were added to an already crowded theater, it is fair to wonder about their marginal impacts despite the institutional pioneering and solid logic behind internalizing forest's benefits.

We use observed rates of deforestation on non-PSA lands to estimate the crucial counterfactual, i.e. what would deforestation on PSA land have been if payments had not been made. For 1997-1999, clearing on all non-PSA land occurred at a rate of 0.21% per year, i.e. well under 1%. In contrast, the ability of Costa Rican conservation authorities to enforce contracts meant that none of the land enrolled in PSA was cleared. If we use the all-non-PSA rate to estimate PSA impact, we expect that over 99% of enrolled parcels would have provided ecoservices without payments.

That suggests the value of targeting. We should add that even if targeting proves difficult, the PSA could be justified if payments are low while environmental benefit per unit impact is high. Even if so, we would propose trying to target to significantly raise benefits and/or save funds.

The 0.21% lower annual deforestation may be a biased estimate of the impact of the payments if the PSA-enrolled parcels differ from non-PSA lands.[1] If the parcels receiving payments faced a

[1] We note that Miranda et al. 2003 and Zbinden and Lee 2005 find PSA participants differ from non-participants.

higher deforestation rate than did other parcels, then the above underestimates the PSA's impact. For instance, if parcels far from cities feature both lower threat and higher costs of participation, those in PSA might face higher threat. Or local conservation expertise might permit targeting of PSA at higher-threat locations (as suggested in Pfaff and Sanchez 2004 e.g.[2]). If so, low average deforestation on non-PSA land would not rule out an impact from the selectively allocated PSA.

The opposite bias could also arise. PSA parcels might face lower threat, for instance if land with higher returns to clearing is more likely to be cleared and is unlikely to participate. If returns to clearing are higher than PSA payments, land is unlikely to be enrolled in this voluntary program. Thus, lower-returns parcels could well be expected to dominate the observed enrollment in PSA.

We find that the PSA parcels do indeed face lower threat. For instance, overlaying 1997-2000 predicted annual deforestation (from a pixel-level regression for the 1986-1997 deforestation) on the map of PSA parcels shows that enrolled parcels faced a lower average deforestation threat. Thus the already low 0.21% annual impact estimate for the PSA is, if anything, an *overestimate*.

To go beyond this bounding to an impact estimate that addresses the bias in the 0.21% value, we employ matching approaches to select a more appropriate base of comparison than all non-PSA. Across many applied empirical fields, matching estimators are becoming increasing popular (see Heckman, Ichimura and Todd 1997 and 1998, Dehejia and Wahba 1999 in *JASA* following many contributions there including Rosenbaum and Rubin 1984, and increasing review of the methods and applications such as in Frolich 2004, Abadie and Imbens 2006a, Morgan and Harding 2006).

We apply two matching estimators to compare the complete lack of deforestation in PSA land with the deforestation in untreated land that is 'similar' to PSA land. The first is a nearest-neighbors propensity score matching estimator (Rosenbaum and Rubin 1983, Hill et al. 2003) that defines similarity based on the land's estimated probability of being enrolled in the PSA. The second is a nearest-neighbors covariate matching estimator (Abadie and Imbens 2006a) that defines similarity using the distances, in matching-covariate space, between treated and matched.

Applying PSM with one untreated match per treated observation usually finds no clearing in the control group as the PSA is a small policy, spatially speaking, so we have only 41 treated points. Increasing the matches from 2:1 up through 50:1 leads the PSA impact estimate to stabilize around 0.08% deforestation per year, or about one third of the estimate from all non-PSA lands.[3]

To test robustness, we apply covariate matching (CM) and then redo the naïve, PSM and CM

[2] From a variety of perspectives, many involving benefits but not costs, and mostly ignoring deforestation threats, authors have suggested reasons for not implementing policies randomly across space but instead optimally targeting (see, for instance, Tubbs and Blackwood 1971, Gehlbach 1975, Williams 1980, Cocks and Baird 1989, Church et al. 1996 and Csuti et al. 1997, Polasky et al. 2000, Camm et al. 2000, and Polasky et al. 2001 for examples and for discussions of the literature). Even when threat is not explicitly involved, such policy allocation processes may enforce positive, negative or zero correlation with threat. Such correlations matter for evaluation of the policies. Robalino et al. 2007 discuss the impact of a change in allocation rule for post-2000 PSA contracts (more below).

[3] This analysis, and thus this result, significantly extends previous suggestions of relatively low impact from PSA (e.g. Sierra and Russman 2006, Harshorn et al. 2005, Sanchez et al. 2007). It contrasts with work by FUNDECOR (presented at WWF by F. Tattenbach, not peer-reviewed as far as we know, but perhaps available upon request).

estimations within regressions with covariates to adjust for biases from matching imperfections. We then test the PSM result for robustness to dropping the observations with only poor matches.

Comparing for six matches per treated observation, the CM estimate is slightly higher than PSM, though closer to it than to the naïve estimate. Bias adjustment brings the naïve estimate down to thee level of the unadjusted CM, while raising the PSM estimate to essentially the same value, but it drives the CM estimate to zero. Importantly, the t-statistics on the CM estimates which should have valid standard errors (Abadie and Imbens 2006a) indicate insignificance of both the unadjusted CM and the bias adjusted CM effects. One takeaway is that match method and bias adjustment certainly can matter yet all matching and adjusted estimates are lower than the naïve.

2. Land-Use Choice with Payment Option

The most basic model of a land owner choosing land use to maximize returns provides a useful framework for communicating several issues that constrain payment impact and its estimation. Figure 1 orders land according to the relatively profitability of clearing, i.e. profits in clearing minus profits in forest, with agriculture more favored to the right. Where the relative profits are greater than zero, the land is predicted to be deforested. With no payments, forest will remain within $[0, x^N]$ while the forest will be cleared from the rest of the land, i.e. to the right of x^N.

If payments for environmental services compete against the gains from non-forest land uses, then landowners sign up for payments only in $[0, x^P]$, where the payment is larger than other options. Not all who wish to apply would modify their behavior as a result of the payment. Those in the interval $[0, x^N]$ would not, as with payments or without they would be forested. In contrast, the parcels in the interval $[x^N, x^P]$ would deforest in the absence of payments but not if being paid.

Thus, the program impact depends upon the fraction of enrollment from $[x^N, x^P]$, denoted by α. If α equals 1, i.e. only land from $[x^N, x^P]$ is enrolled, then all payments prevent deforestation. On the other hand, if α equals 0, i.e. only land from $[0, x^N]$ is enrolled, the payments have no effect.

We will estimate α by finding locations outside the program that are similar to the parcels in the program and computing the deforestation rates for those places. The percentage of those places that was cleared is an estimate of α. If all were cleared, all were from $[x^N, x^P]$ and so α equals 1. Putting that another way, then the PSA payments program would be said to be 100% efficient.

We note that not all who may wish to apply on the basis of being in $[0, x^P]$ will in fact apply, as some land owner may not know about the PSA program or may face high costs of application. Further, not all those who apply are guaranteed to be enrolled. The PSA may not have the funds.

Assuming for the moment that all of $[0, x^P]$ applies, exactly which parcels are enrolled affects the accuracy of simple impact estimates. If $\alpha = 1$, i.e. targeting is good and all of $[x^N, x^P]$ is enrolled but nobody is from $[0, x^N]$ then the forested locations outside the program will be only in $[0, x^N]$. Those are not locations similar to the enrolled. None will be cleared, though all of the enrolled would have been, and α would be underestimated at zero even though all payments had impact. Should $\alpha = 0$, then it would be overestimated, at one, though the payments made no difference.

More generally, accurate estimation of α requires that there exist parcels outside the program that are similar to those enrolled. We believe this is the case, i.e. that parcels which would be cleared with payments and parcels which would not, will be found both in the program and outside of it. This belief is supported by the observation that the PSA was oversubscribed and that, at least for the contracts studied in this paper, the dominant agency selection rule was first-come-first-serve.

3. Data & Methods

3.1 PSA

We obtained information about the PSA program from FONAFIFO. We focus our analysis on forest-conservation contracts that cover most of the hectares enrolled. In 1997, there were over 88,000 hectares under such contracts, making up over 86% of total PSA area. In 1998, more than 47,000 more hectares were enrolled under such contracts, which were over 79% of PSA's total. Finally in 1999, more than 55,000 more hectares were enrolled in forest protection contracts, and these protection contracts again made up over 86% of the total set of hectares enrolled in PSA.

Payments vary according to the year when contracts are signed (noting that they are distributed in equal amounts during the five years that the contract lasts). In colones, payments per hectare were 50,000 in 1997 then 60,000 in 1998 and 1999. On average, that is about $220 per hectare.

We obtained information that depicts forest protection contracts for 1997, 1998 and 1999 across Costa Rica (see Figure 2). The precision of this information allows us to describe the elements that determine the location and to characterize the type of land where the payments took place.

3.2 Forest & Factors In Land Use

We obtained geographic information about the spatial distribution of forest in 1986, 1997 and 2000 from the University of Alberta (Sanchez-Azofeifa et al. 2003). These maps allow us to estimate the amount of forest in each of these years, i.e. annual deforestation rates at the national level as well as what parcels exactly were deforested between 1986 and 1997 or between 1997 and 2000. Kerr and Pfaff 2007, using the same maps, found annual deforestation between 1986 and 1997 to be about 1% and annual deforestation between 1997 and 2000 to be around 0.20%.

Additional maps with the location of rivers, cities, national parks, schools, sawmills, national and local roads and slopes were obtained from the Ministry of Transport and the Instituto Tecnologico. We also use a vegetation description based on Holdridge Life Zone criteria. Costa Rica is covered by 12 zones: humid pre-montane, humid lower-montane, tropical humid, very humid pre-montane, very humid lower montane, very humid montane, tropical dry, pluvial pre-montane, pluvial lower-montane, pluvial montane and paramo. We also use the ministry of agriculture's administrative divisions (Central, Heredia, Huetar Norte, Huetar Atlanctica, Brunca, Pacifico Central and Chorotega) to generate regional dummies are controls within the analyses.

3.3 Units of Analysis

Ten thousand locations were randomly drawn across the 51,000 squared kilometers of Costa

Rican land. We have one location every five square kilometers in average. We eliminate some of these locations from the analysis because clouds covered the land when satellite pictures were taken and/or because, according to experts, the information from the satellite picture was inconclusive (Pfaff and Sanchez 2004). We also eliminate from the sample those observations located within government owned protected areas, which cannot be part of the PSA program.

We focus on forested locations, as forest protection contracts are given only for forested areas. For 1986, there are 1,882 locations covered by private forest. By 1997, there are 1,770 locations in private land covered by forest. By 2000, the number of locations covered by forest is 1,759.

For each location, we find the distances to the closest national road, closest local road, closest sawmill, closest river, closest national park, closest school and the closest already-cleared area. We also find the distance from each location to the country's capital, San Jose, to the two main ports, Limon and Caldera, and to the closest canton's capital. Additionally, we classify locations by life zones[4] and by administrative units. Finally, for our treatment measure we verify which pixel locations are in areas where forest protection contracts were signed in 1997 through 1999. We find 60 points in the PSA parcels, of which none are cleared, and 1710 points outside PSA.

3.4 Matching

3.4.1 What Is 'Similarity' ?

As noted in Section 2, we will estimate α and thus also PSA impact using untreated locations that are 'similar' to parcels receiving payments, by computing deforestation rates for those untreated. However, while those facilitating the application of such techniques agree on this (Rubin 1980, Rosenbaum and Rubin 1983 and many others), similarity has been defined in different ways.

To examine robustness, and to highlight the importance of this definition, we apply two rather different matching estimators. The first is a nearest-neighbors propensity score matching estimator (Rosenbaum and Rubin 1983, Hill et al. 2003*)*, using a fixed number of matched control observations for each treated observation and varying that number from 1 to 50.[5] Propensity score matching estimators define similarity based on estimated probabilities of being treated which are generated by a first-stage regression for whether observations are (not) treated.

The second matching approach we apply is a nearest-neighbors covariate matching estimator (Abadie and Imbens 2006a) using an inverse weighting matrix to account for the difference in the scale of the covariantes. Here again, we employ a fixed number of best matches per treated observation; in Table 2, we apply matching approaches for six matches per treated observation. Covariate matching estimators define similarity without a first-stage regression but rather using the simple distances, in the space of the matching covariates, between the treated and matched.

The computation of standard errors is another difference in how those advancing such techniques

[4] Some small life zones did not appear in the sample
[5] A natural alternative to a fixed number of matched untreated points per treated observation would be to fix the window defining how good a match is required for inclusion and to let the number of matches be endogenous to the quality of the match for each treated observation. This option is related to the discussion in subsection 3.4.2 below.

have applied them. Abadie and Imbens 2006b show that the common practice of bootstrapping standard errors is invalid with non-smooth, nearest-neighbor estimators such as the propensity score matching estimator with a fixed number of matches that we have chosen to present here (contrasted with kernel versions that assign smoothly declining weights to progressively less-well-matched untreated observations). For propensity score matching, then, we do not bootstrap but follow Hill et al. 2003 in calculating weighted standard errors. For lack of certainty, though, we lean more on the covariate matching standard errors which follow Abadie and Imbens 2006a.

We conduct balancing tests for all matching estimators (e.g. Table 5, extending Tables 2 and 3). Balancing compares means of the matching covariates for the treated and matched using a t-test.

3.4.2 Match Quality, OLS & Matching Subsets

For all of the helpful analytics and important choices described above, the matching approaches control solely for selection on observable factors, unlike using instruments for where policies go. Our experience has been that upon recognizing this, many analysts ask why or if OLS is inferior. The basic reasoning is useful *per se* and it motivates further adjustments in matched estimations.

In short, the attempt to match treated with untreated observations explicitly examines whether in fact there exist untreated observations similar (by whatever criterion) to the treated observations. To the extent that the observed characteristics are not similar in these two groups, OLS uses the information at hand to control for differences but the burden on the specification is considerable:

> "Unless the regression equation holds in the region in which observations are lacking, covariance will not remove all the bias, and in practice may remove only a small part or it. Secondly, even if the regression is valid in the no man's land, the standard errors of the adjusted means become large, because the standard error formula in a covariance analysis takes account of the fact that extrapolation is being employed. Consequently the adjusted differences may become insignificant merely because the adjusted comparisons are of low precision. When the groups differ widely in x, these differences imply that the interpretation of an adjusted analysis is speculative rather than soundly based." (Cochran, in Rubin 1984).

This shortfall of OLS where the untreated are not similar to the treated, however, also relates to matching procedures when match quality is not good. That can motivate using a subset of the treated observations, and their matches, in particular dropping treated points with poor matches.

Crump et al. 2006 addresses the issue of a lack of covariate overlap, noting that many common estimators become sensitive to the choice of specification (much as Cochran had noted for OLS, and following also related prior work including Heckman, Ichimura and Todd 1997 and 1998). Crump et al. characterize optimal sub-samples for which treatment effects can be estimated most precisely, which under some conditions can be characterized by a rule based on the propensity score. We do not use these optimal subsamples but do emphasize this point and examine results for robustness to dropping relatively-high-propensity-score treated observations, which we find do not have good matches among the untreated (see Figure 6 and also Cochran and Rubin 1973).

4. Results

4.1 Simplest Naive Estimate of PSA Impact

Ideally, we would obtain the level of rents of each landowner enrolled in the program and sum all those hectares where rents for non-forest uses are higher than rents for forest uses. Lacking this information, as discussed an alternative strategy is to observe the deforestation rate in areas outside the program. We calculate the fraction deforested by 2000 of the points forested in 1997.

Consistent with previous findings (Pfaff and Sanchez-Azofeifa 2004 and Kerr et al.2006), we find (Table 6 first row) that deforestation between 1997 and 2000 in areas outside the program is low. Specifically, the three-year deforestation rate equals 0.63%, i.e. less than 1%. The implied annual deforestation rate during this period is 0.21%, i.e. one fifth of one percent.[6] According to this estimate, very little of the land enrolled in PSA was actually protected from deforestation (e.g., we would compute that 5-year contracts postponed deforestation in 928 of the eighty-eight thousand hectares enrolled in 1997, with similar fractions applying to 1998 and 1999 contracts[7]).

4.2 Non-random Enrollment Bias

Longstanding conservation expertise in Costa Rica, as well as the structure of the PSA program, suggest that enrollment may be non-random. Costa Rican experts could target the PSA so as to have real impact. On the other hand, coupled with the voluntary nature of the PSA program, variation across landowners in relative profits from clearing could lead PSA enrollment to be dominated by parcels so poor for agriculture, e.g., that they would never be cleared in any case.

One way to directly examine this direction of actual net bias is the use of predicted deforestation. We run a pixel regression to generate probabilities of 1997-2000 deforestation (see appendix for the regression, which follows upon work in Robalino and Pfaff 2007 and Kerr and Pfaff 2007).

With the probabilities in hand, we can examine whether enrollment leans towards higher or lower deforestation threats. Following the logic in Section 2 and Appendix 1, if it leans towards lower pressure then our already low initial impact estimate in 4.1 is an overestimate of impact.

To examine this, we divide locations by rent and calculate the fraction enrolled from each group. Using five categories, Table 1 shows that when estimated threats that are positively correlated with agricultural profitability increase, the fraction of parcels that enroll in the PSA decreases. We conclude that 0.21% annual deforestation rate in untreated areas overestimates PSA impact.

Table 1 – PSA enrollment bias towards lower threat

---

[6] Another way to communicate this result is that 99.79% of PSA funds did not change behavior during the first year of the contract. During the second year an additional 0.21% of the forest stock is saved, meaning 0.42% of the original stock has been protected and thus 99.58% of funds have no impact. The same logic implies 99.37%, 99.16% and 98.95% "waste" for the final three years. Averaging across the five years, this upper bound on the impact of PSA is 0.63%. On this basis, the results suggest that over 99% of the PSA funds allocations did not change land use.

[7] Given total PSA expenditure on the order of 10 million US dollars per year for these contracts and this estimate of hectares on which deforestation is postponed by each year's contracts on the order of 1000, another way to discuss the efficiency of such payments is a very rough estimate of around 10,000 US dollars per hectare-year protected. Since such an estimate takes into account the fraction of enrolled forest that we estimate was not going to be cleared in the absence of payments, clearly it is much higher than the cost per unit service suggested by official documents.

Probability of Deforestation (Bins)	Average Probability	Fraction Enrolled	Number of Observations
0.00-0.15	0.05	0.040	1163
0.15-0.30	0.21	0.028	348
0.30-0.45	0.36	0.010	100
0.45-0.60	0.51	0.000	23
0.60-0.75	0.61	0.000	3

4.3 Matching Analyses

Within the 1,770 privately-held forest locations in 1997, i.e. within the relevant subset of our sample of 10,000 pixel locations for which we have values for relevant variables, only 41 are in PSA polygons. For estimates applying either propensity score matching or covariate matching approaches, we will use some multiple of 41 points from the 1729 non-PSA forested locations.

4.3.1 Propensity Score Matching

We run a probit (Appendix 3) to determine the effects of a location's observed characteristics on its probability of being treated. We predict the probability of being treated for all 1,770 points, then for each PSA point we find the non-PSA location(s) with the closest predicted probabilities.

First we match a single non-PSA point to each PSA location. This yields 41 controls (Figure 4 maps them in connection with their treated points). In fact, in the match presented in Figures 3 and 4 and Table 2, we use 39 distinct control points. Two untreated points are matched to more than one treated point. Re-matching is more frequent when using more than one control point per treated point and it changes the appropriate standard errors (something not yet reflected here).

Figure 3 and Table 2 convey the basic point of the matching approach. Figure 3 shows the result of matching by predicted probability of treatment. The distribution of predicted probabilities for the matched untreated (in red) better approximates the treated (in green) than does all untreated (in blue). Thus the matched untreated form a better control group. The right-most box, putting the three groups on the same numerical scale, shows that the PSA is small enough that within the untreated there are many points to be matched to the treated, in each treatment probability bin, even though as just noted the untreated are of course on average less likely to have been treated.

Table 2 compares the treated, all-untreated and matched-treated groups in terms of the means of the covariates used for doing the matching, i.e. explanatory factors in the treatment regression. Balancing via the propensity-to-be-treated ideally leads the matched untreated group to be more like the treated group, for each of these variables, than is the group of all the untreated. While in two cases that is weakly not true here, on the whole the matching seems to aid in the comparison.

Tables 3 and 4 demonstrate the potential for implementing 'filters' within the matching process (to some extent following the discussion above of Cochran and Rubin 1973, Heckman et al. 1997 and 1998 and also Crump 2006, although see Figure 6 and Table 5 for more along those lines). Here we might be concerned that the fact that PSA payments were made in an area has an impact on the land use in neighboring areas (Robalino and Pfaff in progress finds such effects for parks, as does Andam et al. 2007, which notes this has implications for the appropriate control group). Restricting matches to be at least 10 kilometers from the treated point, to avoid such influence, leads to two changes in match (Table 4 in bold) but little impact upon the balancing (Table 3).

As previewed in the introduction, since the PSA is such a spatially small policy Tables 3 and 4, which employ only a single match for each of the only 41 treated locations, generate an estimate of zero for the impact of the PSA program. Zero deforested locations appear within the set of 41 matched untreated points. Since we know that there is positive although low deforestation during this period, these estimates appear to derive from the use of only a single match per treated point.

To address this possibility, we redid the matching procedure expanding the set of matched points to two matched untreated controls per treated point, then 3:1, 4:1, 5:1 and up through 50:1. As seen in Figure 5, the PSA impact estimate rises above zero as a positive number of cleared points is included in the control set (and the integer nature of this leads to jumps whenever another deforested point is included). As the number of the matched untreated control points used rises, the estimated impact stabilizes around 0.08% deforested annually, or 0.24% cleared in the three year 1997-2000 period (about one third of the 0.64% from all untreated seen in Tables 2 and 3). The right-hand part of Figure 5 shows that of the 50 estimates, the majority of the three-year impacts fall between 0.2% and 0.3%, i.e. less than one percent and less than the biased estimate.

Figure 6 then raises the possibility of a different kind of filter within the matching process. It shows that for relatively high propensity treated points, i.e. treated points whose characteristics made it likely that they would be treated, there are no very good matches to untreated points. The poor matches lessen the value of matching, so we test for robustness to dropping these points.

4.3.2 Comparing Estimators

To test robustness, we apply covariate matching (CM) and then redo the naïve, PSM and CM estimations within regressions with covariates to adjust for biases from matching imperfections (see, for instance, Rubin and Thomas 2000). We then test the PSM results for robustness to dropping the observations which have only poor matches (which are indicated well in Figure 6).

Table 5 conveys the basic point of the matching, for Table 6's basic PSM match (like Table 2 but for six matches per treated and with tests). To address the key point directly, the P-values of comparisons between the means of matching covariates for the treated and the matched untreated are almost all quite high, suggesting little evidence of significant differences in the covariate sets.

In Table 6, for six matches per treated observation, the CM estimate is slightly higher than PSM, though closer to it than to the naïve estimate. Bias adjustment brings the naïve estimate down to thee level of the unadjusted CM, while raising the PSM estimate to essentially the same value, but it drives the CM estimate to zero. Importantly, the t-statistics on the CM estimates which should have valid standard errors (Abadie and Imbens 2006a) indicate insignificance of both the unadjusted CM and the bias adjusted CM effects. One takeaway is that match method and bias adjustment certainly can matter yet all matching and adjusted estimates are lower than the naïve.

5. Discussion

We find a very small impact on deforestation from payments for environmental services in Costa Rica. Our initial estimate was that the five-year forest-protection contracts originated in 1997, 1998 and 1999 prevented deforestation in their first contract years on only 0.21% of the land enrolled (implying that only 1% of PSA area was actually saved during the fifth contract years).

We then used spatially-varying pixel-level predicted deforestation rates to show that if anything that is an *overestimate* of the program's impact. The likely reason is that PSA is a first-come, first-served program for which households volunteer. With per-hectare payments fixed across the country, one might expect that those willing to forego production and enroll in the PSA had relatively low returns. It appears that most had negative net returns to deforestation, i.e. were not going to clear their land anyway given all of the factors in clearing decisions other than the PSA.

This result would appear to have analogs relevant for the design of both the international and the domestic elements of Clean Development Mechanism components of the Kyoto Protocol or its possible successors. As Montero has showed, if baseline are not known with certainty by those implementing a policy, much as for the PSA self-selection into the policy could lower efficiency. Further, even if the issues with national-level baselines are acceptable at the international level, in internally implementing international agreements countries may well try PSA-like payments.

To go beyond that bounding argument, we applied matching procedures to construct a control or comparison group of untreated areas more 'similar' to the PSA locations than is the entire set of untreated areas. Using propensity score matching, as we added more matched points per treated observation our impact estimate stabilized around 0.08% clearing per year, i.e. at about one third of the estimate derived using all untreated points. Robustness checks adding covariate matching and then adding bias adjustment to each of the methods confirm that the initial estimate is too high. The covariate matching standard errors suggest that there is no significant impact of PSA.

While some prior work also suggests low impacts of PSA (Hartshorn et al. 2005, Sierra and Russman 2006, Sanchez et al. 2007), work by FUNDECOR claims large impacts and it has been noted that different researchers have used different time periods and different scales of analysis including different data aggregations. To help address whether such differences may explain the possible disagreement, Robalino et al. 2007 apply both the approach in this paper and that of Sanchez et al. 2007, which used 5x5k grid cells and did not address the dissimilarity of PSA locations, to the post-2000 PSA data. The estimate of less than a 1% annual impact is repeated.

Yet there are differences. First, despite net reforestation during 2000-2005 a greater forest area was cleared than during 1997-2000, indicating more opportunity for a policy to have an impact, and indeed the impact estimate is higher (more like half a percent per year, versus under a fifth). Second, targeting changed, not to consciously focus on forest impact but instead to aim at places with higher environmental benefits. It appears that the seemingly unintended consequence of this was a reduction in the influence of the voluntary element of the PSA program, in that the biases found here towards the lower-threat locations do not appear to be present in the post-2000 PSA.

It remains important to emphasize the institutionally innovative nature of the PSA policy. Further, simple adjustments could increase its impact. In addition, effectiveness should be measured against all stated goals, one of which is distribution, i.e. compensating owners for the benefits to society their land uses provide (FONAFIFO, 2005). However, as the guiding principle for distributing the surplus has not been made explicit, as far as we know, it is unclear at least to us what payments level, or distribution, would be suggested by such distributional rationales.

A focus on higher-pressure parcels could in principle raise impact. First, if the program remains oversubscribed, in selecting parcels to be admitted the agency can target those who would clear in the absence of payments, rejecting those who would not clear anyway. Targeting does appear to be increasing within the PSA, although this includes along dimensions of environmental services benefits which may not be positively correlated with rents. Second, the program design could be adjusted to permit higher payments to areas under higher threat of clearing due to their higher returns. One payment five times as high as the current one could yield a positive impact even if five payments under the current design yield nothing. These possible adjustments require the agency to have information about rents. That is not likely to be perfect but enough may exist.

Such adjustments could change the PSA's distributional effects. For a fixed budget, having high payments means concentrating payments on fewer people, e.g. providing nothing to four of any five people who would have enrolled in the current regime but five times as much as currently paid for the fifth person. Such a shift could in principle help to alleviate poverty if that one person is relatively poor. Targeting higher rent land, though, may mean that the person receiving payments has relatively high profits and relatively high wealth. Combining the targeting of the higher rents with fewer higher payments could, then, shift the program away from redistribution.

One final perspective on this policy is that in the case of Costa Rica, PSA impact could be linked to the forestry law which appears to have essentially guaranteed that PSA could not have a large direct impact on deforestation. As noted, Forestry Law 7575 raised significant obstacles for any post-1997 deforestation. It was essentially a taking, removing a right to clear land (noting that Costa Rica had previously debated issues about the compensation for takings of land, even for environmental reasons including per the establishment of the parks on previously private land).

It is possible that establishing the PSA could have made possible the passing (or acceptance) of such a law, which appears to have been respected, something we suspect would not be possible in most other countries. If that is the case, then the existence of the payments has had impact on deforestation. Further, the exact distribution of the payments might in principle not matter very much for this impact, although we would think those who wanted to clear would have lobbied hardest to get payments, i.e. the same types of landowners who do not seem to be in the PSA. Whatever the case for Costa Rica, we suspect this is not a model of impact others will copy.

References

Abadie, A. and G. W. Imbens (2006a) "Large sample properties of matching estimators for average treatment effects", *Econometrica* 74 (1): 235-267

Abadie, A. & Imbens, G. W. (2006b) "On the failure of the bootstrap for matching estimators." NBER Technical Working Paper #325, http://www.nber.org/papers/t0325.pdf.

Andam, K., P. Ferraro, A. Pfaff, J. Robalino and A. Sanches (2007). "Evaluating Policies to Secure the Provision of Ecosystem Services: an econometric analysis of protected areas". Mimeo, Department of Economics, Georgia State University.

Bruner, A. G., Gullison, R. E., Rice, R. E. and da Fonseca, G. A. B. (2001) "Effectiveness of parks in protecting tropical biodiversity." *Science* 291: 125-128.

Camm, J.D., S.K. Norman, S. Polasky, and A.R. Solow, 2002, Nature Reserve Selection to Maximize Expected Species Coverage, Operations Research 50(6), 946-955.

Chomitz, K., Brenes, E. and Constantino, L. 1998. Financing Environmental Services: The Costa Rican Experience. 10, Central America Country Management Unit, Latin American and the Caribbean Region, The World Bank

Church, Richard L., David M. Storms and Frank W. Davis, 1996, Reserve Selection as a Maximal Covering Location Problem, Biological Conservation 76, 105-112.

Cocks, K.D. and I.A Baird, 1989, Using Mathematical Programming to Address the Multiple Reserve Selection Problem: An Example from the Eyre Peninsula, South Australia, Biological Conservation 49, 113-130.

Cochran, W. and D. Rubin (1973). "Controlling Bias in Observational Studies: A Review". *Sankhya*, Series A 35: 417-446.

Crump, R.K., V.J. Hotz, G.W. Imbens, O.A. Mitnik (2006). "Moving The Goalposts: addressing limited overlap in estimation of average treatment effects by changing the estimand". Mimeo, U. California Berkeley, Dept. of Economics (supplemental proofs on Imbens' web site).

Csuti, B., S. Polasky, P.H. Williams, R.L. Pressey, J.D. Camm, M. Kershaw, A.R. Kiester, B. Downs, R. Hamilton, M. Huso and K. Stahr, 1997, A Comparison of Reserve Selection Algorithms Using Data on Terrestrial Vertebrates in Oregon, Biological Conservation 80, 83-97.

de Camino Velozo, R. and World Bank. 2000. *Costa Rica : forest strategy and the evolution of land use*. Washington, DC: World Bank

Dehejia, R. and S. Wahba (1999). "Causal Effects in Nonexperimental Studies: re-evaluating the evaluation of training programs". *J. of the American Statistical Association* 94: 1053-1062.

Echevarria, M. and International Institute for Environment and Development. Environmental Economics Programme. 2004. *The impacts of payments for watershed services in Ecuador emerging lessons from Pimampiro and Cuenca*. London: International Institute for Environment and Development

Ferraro, P. 2001. “Global habitat protection: limitations of development interventions and a role for conservation performance payments” *Conservation Biology* 15(4):990-1000.

Ferraro, P. J. & Pattanayak, S. K. (2006) Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology* **4**(4): e105.

FONAFIFO (2006) Servicios Ambientales: Estadisticas PSA http://www.fonafifo.com/text_files/servicios_ambientales/distrib_ha_Contratadas.pdf

Frank, G and F. Muller (2003), "Voluntary approaches in protection of forests in Australia" Environmental Science and Policy. 6 (3) 261-269

Frolich, M. (2004) Finite-sample properties of propensity-score matching and weighting estimators. *The Review of Economics and Statistics* **86**(1): 77-90.

Gehlbach, F.R., 1975, Investigation, evaluation, and priority ranking of natural areas, Biological Conservation 8, 79-88.

Hartshorn, G., P. Ferraro, B. Spergel and E. Sills (2005). “Evaluation of the World Bank – GEF Ecomarkets Project in Costa Rica”. Report of ‘Blue Ribbon Evaluation Panel’.

Heckman, J.J., H. Ichimura and P. Todd (1997). “Matching As An Econometric Evaluation Estimator: evidence from evaluating a job training program”. *Review of Economic Studies* 64: 605-54.

Heckman, J.J., H. Ichimura and P. Todd (1998). “Matching As An Econometric Evaluation Estimator”. *Review of Economic Studies* 65: 261-94.

Hill, J., Waldfogel, J., and J. Brooks-Gunn. (2003) “Sustained Effects of High Participation in an Early Intervention for Low-Birth-Weight Premature Infants” *Developmental Psychology*, 39(4): 730--44.

Kerr, S., and Pfaff, A. 2007. Development and Deforestation: Evidence from Costa Rica. Mimeo, Duke University

Landell-Mills, N., Porras, I. T. and International Institute for Environment and Development. 2002. *Silver bullet or fools' gold : developing markets for forest environmental services and the poor Silver Bullet or Fools' Gold*. Stevenage, Hertfordshire: International Institute for Environment and Development

Miranda, M., Porras, I. T. and Moreno, M. L. 2003. *The social impacts of payments for* *environmental services in Costa Rica : a quantitative field survey and analysis of the Virilla watershed*. London: IIED Environmental Economics Program

Morgan S. and D.J. Harding (2006). “Matching Estimators of Causal Effects: prospects and pitfalls in theory and practices”. *Sociological Methods and Research* 35:3-60.

Pagiola, S. 2002. *Paying for Water Services in Central America: learning from Costa Rica* in S. Pagiola, J. Bishop and N. Landell-Mills, eds., *Selling forest environmental services : market-based mechanisms for conservation and development*. London Sterling, VA: Earthscan

Pfaff, A. S. P. and Sanchez-Azofeifa, G. A. 2004. Deforestation pressure and biological reserve planning: a conceptual approach and an illustrative application for Costa Rica, *Resource and Energy Economics* 26 (2): 237-254

Polasky S., J.D. Camm, A.R. Solow, B. Csuti, D. White and R. Ding, 2000, Choosing reserve networks with incomplete species information, Biological Conservation 94, 1-10.

Polasky, S., J.D. Camm, and B. Garber-Yeats, 2001, Selecting Biological Reserves Cost-Effectively: An Application to Terrestrial Vertebrate Conservation in Oregon, Land Economics 77(1), 68-78.

Robalino, J., A. Pfaff, A. Sanchez, F. Alpizar, C. Leon, and C.M. Rodriguez (2007). “Deforestation Impacts of Environmental Services Payments: Costa Rica’s PSA Program 2000-2005”. Mimeo, CATIE, submitted to special issue on the Costa Rican payments experience.

Rojas, M., Aylward, B. and International Institute for Environment and Development. Environmental Economics Programme. 2003. *What are we learning from experiences with markets for environmental services in Costa Rica? A review and critique of the literature*. London: International Institute for Environment and Development

Rosales, R. M. P. and International Institute for Environment and Development (2003). Environmental Economics Programme. 2003. *Developing pro-poor markets for environmental services in the Philippines*. London: International Institute for Environment and Development

Rosenbaum, P. R. and D. B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika* 70(1): 41-55

Rosenbaum, P. R. and D. B. Rubin (1984), “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score”, *J. of the American Statistical Association.*

Rubin, D. B. (1980). “Bias reduction using Mahalanobis-metric matching.” *Biometrics* 36: 293-298.

Rubin, D. 1984. “William G. Cochran's Contributions to the Design, Analysis, and Evaluation of Observational Studies.” in Rubin, D. (Ed.) Matched Sampling for Causal Effects 2006 Cambridge University Press 7-29

Rubin, D.B. and N. Thomas (2000). "Combining Propensity Score Matching With Additional Adjustments for Prognostic Covariates". J. of the American Statistical Association 95: 573-85.

Sanchez-Azofeifa, G. A. Daily, G. C., Pfaff, A. S. P. and Busch, C. 2003. Integrity and isolation of Costa Rica's national parks and biological reserves: examining the dynamics of land-cover change, *Biological Conservation* 109 (1): 123-135

Sánchez-Azofeifa, A, A. Pfaff, J. Robalino, and J. Boomhower (2007). "Costa Rican Payment for Environmental Services Program: Intention, Implementation and Impact". *Conservation Biology* In Press

Sierra, R, and E. Russman (2006) "On the efficiency of the environmental service payments: A forest conservation assessment in the Osa Peninsula, Costa Rica" *Ecological Economics* 59: 131-141

Smith, R. B. W. 1995. The Conservation Reserve Program as a Least-Cost Land Retirement Mechanism, *American Journal of Agricultural Economics* 77 (1): 93-105

Szentandrasi, S., Polasky, S., Berrens, R. and Leonard, J. 1995. Conserving Biological Diversity and the Conservation Reserve Program, *Growth and Change* 26 (3): 383-404

Tikka, P (2003), "Conservation contracts in habitat protection in southern Finland" Environmental Science and Policy 6 (3) 271-278

Tubbs, C.R. and J.W. Blackwood, 1971, Ecological evaluation of land for planning purposes, Biological Conservation 3, 169-172.

Williams, G., 1980, An index for the ranking of wildfowl habitats, as applied to eleven sites in West Surrey, England, Biological Conservation 18, 93-99.

Zbinden, S. and Lee, D. R. 2005. Paying for environmental services: an analysis of participation in Costa Rica's PSA program, *World Development* 33 (2, Special Issue): 255-72

Acknowledgements This paper extends the Costa Rica Carbon Sequestration project and owes a great deal to that project's members, and in particular Suzi Kerr. We gratefully acknowledge project support from the Canadian SSHRC (Social Sciences and Humanties Research Council), the National Science Foundation, the Tinker Foundation, the Harvard Institute for International Development, the NSF's National Center for Ecological Analysis and Synthesis (NCEAS) at UCSB, and the Center for Environmental Research and Conservation at Columbia University.

Appendix 1 – land-use choice with payment option adding shifts in returns & agency enrollment choice

Extending Figure 1 (see the amended figure just below Figure 1) and the discussion in Section 2, R_t represents the relative profitability of clearing at time *t* and in the figure R increases with *t*. In addition, the payment changes over time, applied only in period *t*, i.e. not in either *t-1* or *t+1*. This makes very clear that the payment only postpones deforestation, as the forest stock will fall over time from x^N_{t-1} to x^P_t to x^N_{t+1} with payments versus from x^N_{t-1} to x^N_t to x^N_{t+1} without them.

Enrollment depends on which landowners wish to participate and how an agency selects among them. Such selection by the agency may be required due to oversubscription of the program, i.e. a lack of sufficient funds to compensate all of those who are willing to participate at period t. Any application or transaction costs for being in such a program could affect enrollment as well.

For budget reasons (relevant for Costa Rica), an agency could randomly enroll a fraction $r < 1$ of those willing to participate. This would make the impact of the policy $r\,(x^P_t - x^N_t)$. The fraction of owners who would change behavior due to the policy is the same as for complete enrollment, i.e. it is $(x^P_t - x^N_t) / x^P_t$. A first approach to estimating impact is to examine the rate of deforestation outside of the program. Under this random enrollment policy, from *t-1* to *t*, that rate will be:

$$q_{t-1,t} \equiv \frac{(x^*_{t-1} - x^p_t) + (1-r)(x^p_t - x^*_t)}{(x^*_{t-1} - x^p_t) + (1-r)x^p_t},$$

i.e. 100% when rents are high enough to not apply plus the $[(x^p_t - x^*_t) \; / \; x^p_t]$ rate for those willing to enroll but randomly not enrolled. We note that because of the shift in rents that leads owners of some previously forested land to not want to apply, this $q_{t-1,t}$ is higher than $[(x^p_t - x^*_t) \; / \; x^p_t]$ and overestimates the share of those in the program whose behavior is changed by the payments. Under random enrollment the deforestation outside of the program is an upper bound on impact.

If the agency has signals of rents, it can target those who would change behavior with payments. If only landowners who would change behavior are enrolled, as noted in Section 2 the 100% rate of impact will be underestimated by non-program deforestation. Summarizing from Section 2 and above, non-program deforestation could over- or under-estimate impact, depending on which land is enrolled. In particular, we now see that random enrollment could yield an overestimate.

Assume the net effect of all factors affecting enrollment is that a fraction r_L of owners in $[0,x^N_t]$ and a fraction r_H from $[x^N_t, x^P_t]$ are enrolled. Then non-the program clearing rate from *t-1* to *t* is:

$$q_{t-1,t} = \frac{(x^*_{t-1} - x^p_t) + (1-r_H)(x^p_t - x^*_t)}{(x^*_{t-1} - x^p_t) + (1-r_L)x^*_t + (1-r_H)(x^p_t - x^*_t)}.$$

As $r_L = r_H$ reproduces random enrollment, in which the non-program deforestation overestimates the program's impact, it is clear that the non-program rate is again an overestimate if $r_H <= r_L$.

Appendix 2 – pixel regression to predict 1997-2000 deforestation

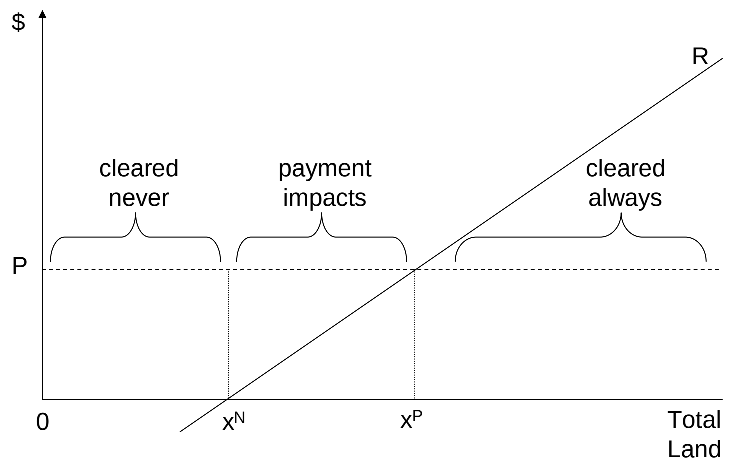
Variable	Coefficient	t-statistic
Life Zone 1	-0.61572	-2.05
Life Zone 3	-0.09405	-0.70
Life Zone 5	-0.09397	-0.33
Life Zone 7	-0.29713	-2.30
Life Zone 9	-0.01409	-0.06
Life Zone 10	-0.10258	-0.29
Life Zone 11	0.387917	0.72
Distance to San Jose	0.00154	0.41
Distance to Limon	-0.00729	-3.49
Distance to Caldera	-0.0043	-1.57
Distance to Local Roads	0.015358	0.63
Distance to National Roads	0.021859	1.39
Distance to to Sawmills	-0.00644	-1.19
Distance to Schools	0.005478	0.93
Distance to Cleared Areas	-1.86565	-6.51
Distance to Main Towns	0.0004	0.08
Terrain Slope	-0.04564	-5.64
Density of National Roads	0.007553	2.64
Density of Local Roads	0.000525	0.27
Sawmills in a 10km radius	-0.03772	-0.88
Main towns in a 10km radius	-0.20131	-1.72
Schools in a 10km radius	-0.0168	-0.55
Cleared Percentage in a 10km radius	0.904865	1.16
(Cleared Percentage in a 10km radius)^2	-1.0067	-1.42
Constant	0.442469	1.06
Observations	1882	
Log-Likelihood	-594.271	

Appendix 3 – propensity score first stage (treatment) regression

```
Probit Maximum Likelihood Estimates

Dependent Variable =        Treatment

McFadden R-squared     =    0.1271
Estrella R-squared     =    0.0295

LR-ratio, 2*(Lu-Lr)    =   49.5524
LR p-value             =    0.0011
Log-Likelihood         = -170.1169
Convergence criterion  =    4.385991e-007

Nobs, Nvars            =   1770,    24
# of 0's, # of 1's     =   1729,    41

****************************************************************
```

Variable	Coefficient	t-statistic	t-probability
R1	0.247046	0.580439	0.561694
R2	0.070003	0.151004	0.879990
R3	-3.864933	-0.043692	0.965155
R4	-0.009696	-0.025662	0.979529
R5	0.168742	0.571188	0.567945
R7	-0.336118	-0.730122	0.465414
DSJ2	0.000029	0.787485	0.431105
DSJ	-0.007757	-0.906528	0.364782
DNR2	-0.008273	-1.541986	0.123258
DNR	0.115253	1.647330	0.099670
DLR2	-0.007436	-0.669165	0.503479
DLR	0.105325	1.107709	0.268140
DNP2	0.000477	1.022719	0.306582
DNP	-0.018201	-0.928828	0.353107
DRW2	-0.005349	-0.509083	0.610758
DRW	0.038219	0.394236	0.693455
PTC2	-0.402397	-1.152992	0.249071
PTC	0.919317	1.759986	0.078585
SDA	-0.023632	-1.720969	0.085434
REL2	-0.748389	-1.939138	0.052645
REL	1.280025	1.919077	0.055137
RAI2	0.115036	2.010769	0.044503
RAI	-0.889616	-1.946670	0.051734
C	-0.641864	-0.653489	0.513527

Figure 1 – land-use choice with payment option



for Appendix 1 – adding changes in returns over time

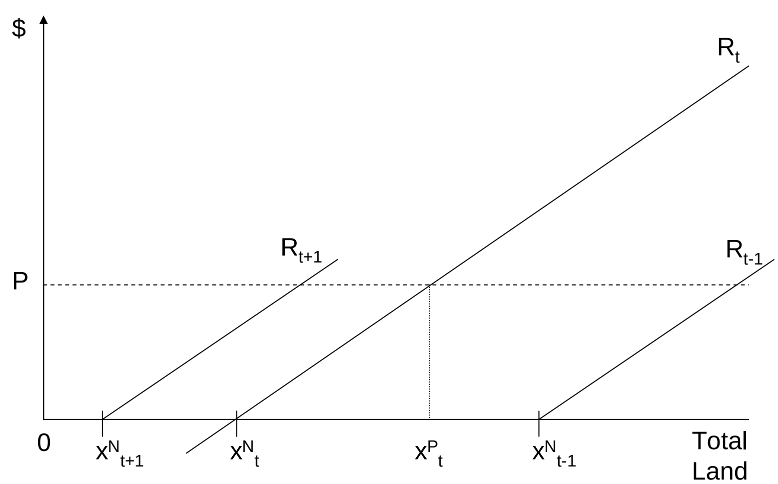


Figure 2 – locations of PSA forest protection contracts

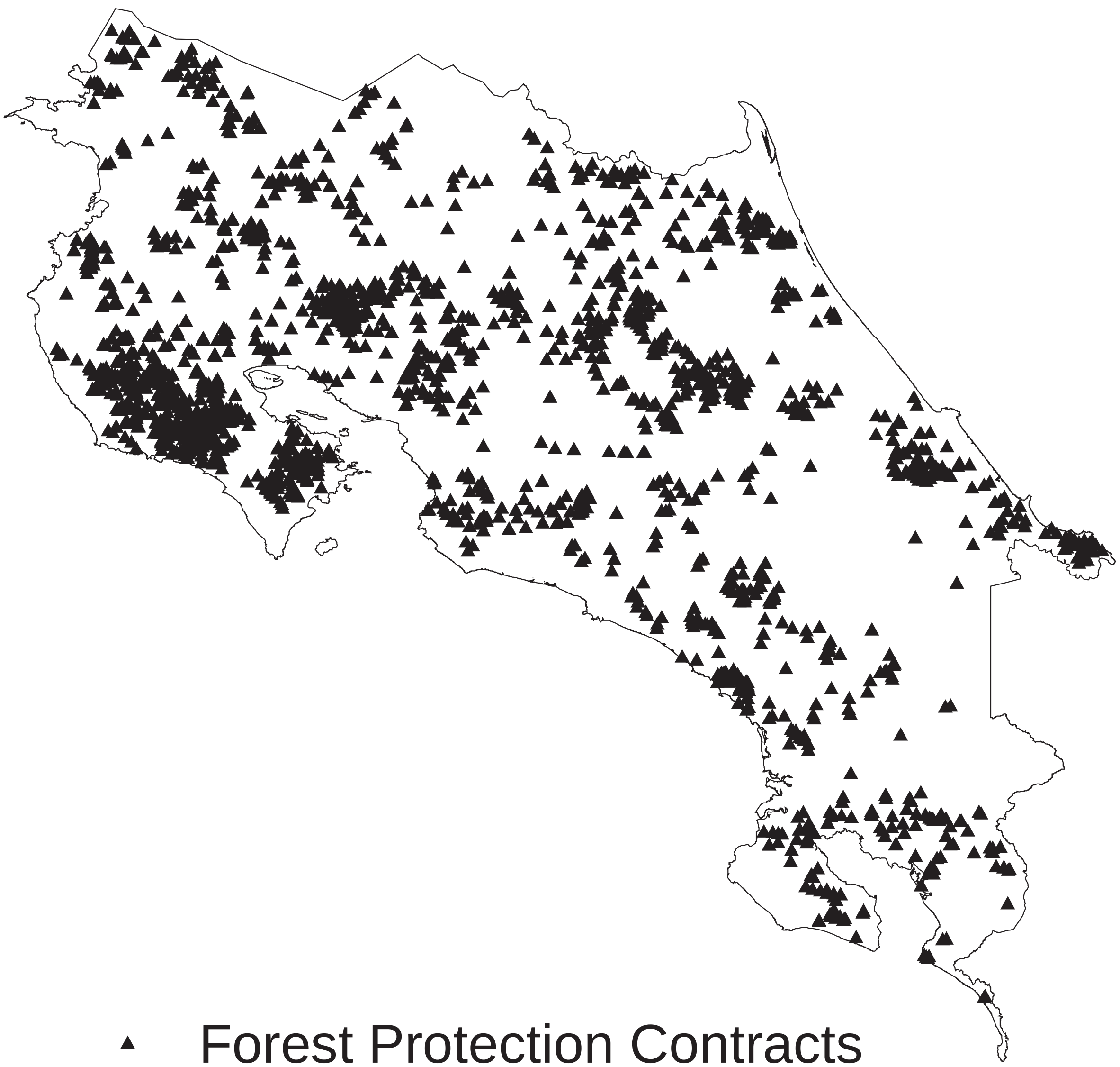


Figure 3 – comparing treated to untreated to matched untreated (PSM, 1 match)

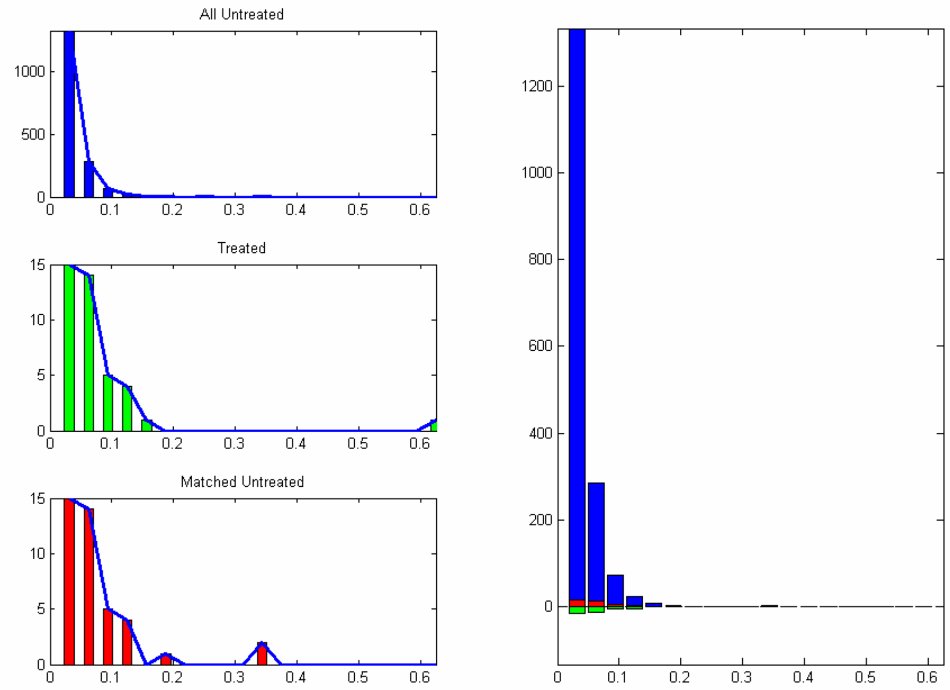


Figure 4 – connecting treated to matched untreated (PSM, 1 match)

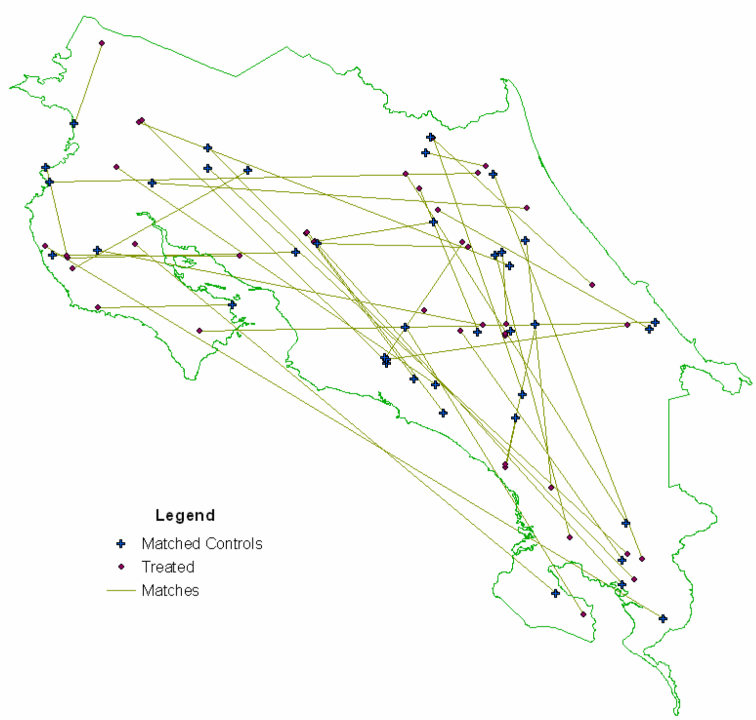


Figure 5 – evolution of PSM impact estimate as the number of matched points rises

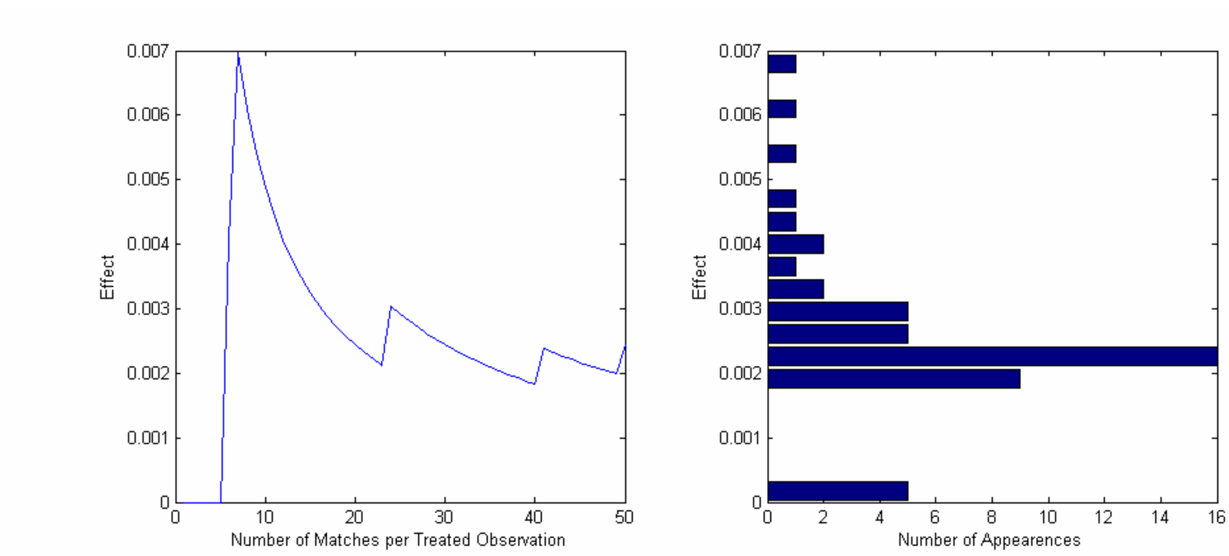


Figure 6 – evaluating PSM match quality as the treated propensity score rises

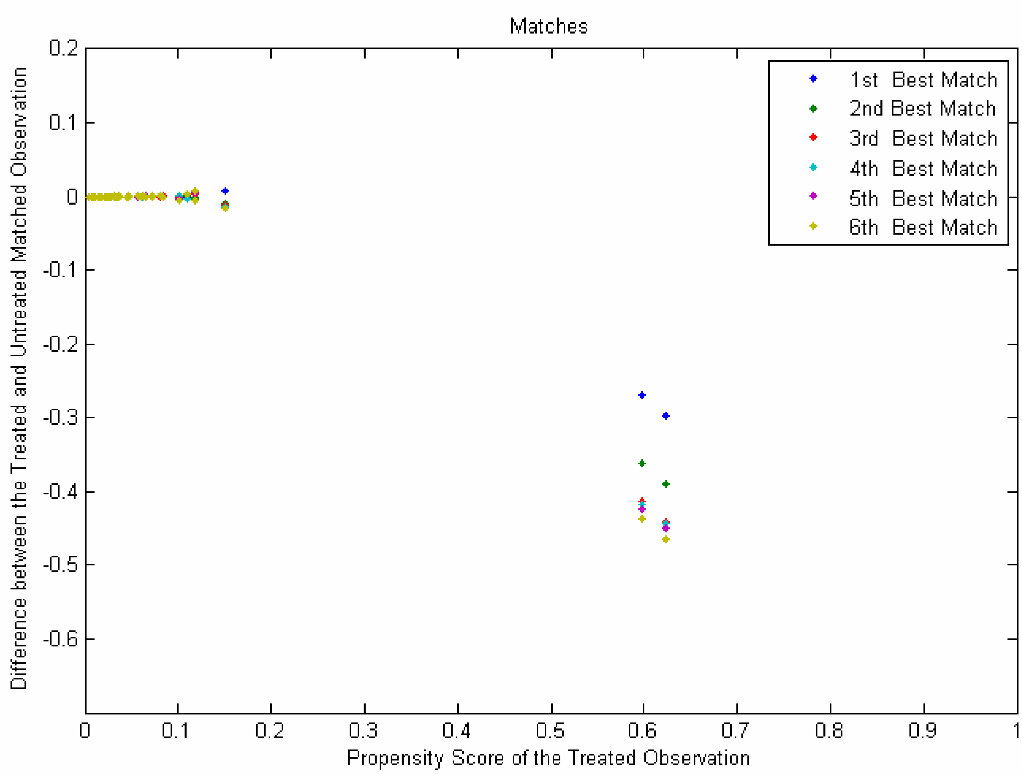


Table 2 -- variable balances (propensity score, 1 match, no restrictions)

Variable	Treated	All Untreated	Matched Untreated
Y	0.0000	0.0064	0.0000
R1	0.2439	0.0943	0.2927
R2	0.0976	0.0821	0.0732
R3	0.0000	0.1533	0.0000
R4	0.1707	0.1643	0.1707
R5	0.1951	0.1805	0.2195
R7	0.0488	0.0873	0.0244
DSJ2	11738.9080	12686.1231	10177.3416
DSJ	92.5821	100.6826	83.5129
DNR2	27.3170	32.0941	29.7573
DNR	4.3823	4.1390	4.5400
DLR2	12.1030	10.5939	11.0984
DLR	2.8501	2.4509	2.7006
DNP2	304.6311	464.0022	324.9706
DNP	14.0342	17.4462	15.4701
DRW2	15.7378	19.0356	19.8404
DRW	3.3053	3.4062	3.6428
PTC2	0.3327	0.1111	0.1679
PTC	0.3662	0.2035	0.2917
SDA	7.1184	6.4256	8.2582
REL2	0.4555	0.4045	0.6039
REL	0.5175	0.4113	0.6088
RAI2	15.4878	12.6653	14.7917
RAI	3.7073	3.4335	3.6439
C	1.0000	1.0000	1.0000

Table 3 -- variable balances (propensity score, 1 match, > 10k restriction)

Variable	Treated	All Untreated	Matched Untreated
Y	0.0000	0.0064	0.0000
R1	0.2439	0.0943	0.2439
R2	0.0976	0.0821	0.1220
R3	0.0000	0.1533	0.0000
R4	0.1707	0.1643	0.1707
R5	0.1951	0.1805	0.2195
R7	0.0488	0.0873	0.0244
DSJ2	11738.9080	12686.1231	10156.2876
DSJ	92.5821	100.6826	83.1992
DNR2	27.3170	32.0941	32.7264
DNR	4.3823	4.1390	4.7645
DLR2	12.1030	10.5939	12.4529
DLR	2.8501	2.4509	2.8114
DNP2	304.6311	464.0022	317.7196
DNP	14.0342	17.4462	14.8786
DRW2	15.7378	19.0356	19.0569
DRW	3.3053	3.4062	3.5670
PTC2	0.3327	0.1111	0.1615
PTC	0.3662	0.2035	0.2848
SDA	7.1184	6.4256	7.9210
REL2	0.4555	0.4045	0.5597
REL	0.5175	0.4113	0.5820
RAI2	15.4878	12.6653	13.9502
RAI	3.7073	3.4335	3.5707
C	1.0000	1.0000	1.0000

Table 4

Distances between Matched Observations
(propensity score approach, 1 match)

No Restrictions	10k+ Restriction
1.9965e+005	1.9965e+005
42178	42178
1.9419e+005	1.9419e+005
3.617e+005	3.617e+005
1.2224e+005	1.2224e+005
2.0714e+005	2.0714e+005
30811	30811
2.1463e+005	2.1463e+005
77698	77698
1.88e+005	1.88e+005
1.3686e+005	1.3686e+005
1.7301e+005	1.7301e+005
90750	90750
2.3757e+005	2.3757e+005
60882	60882
1.0793e+005	1.0793e+005
93702	93702
1.0968e+005	1.0968e+005
75583	75583
69598	69598
2.7418e+005	2.7418e+005
1.0113e+005	1.0113e+005
67477	67477
1.1447e+005	1.1447e+005
2.2893e+005	2.2893e+005
45653	45653
52317	52317
1.2162e+005	1.2162e+005
36375	36375
1.4069e+005	1.4069e+005
1.9673e+005	1.9673e+005
3800	**72054**
3294.7	**71425**
25728	25728
72147	72147
1.2469e+005	1.2469e+005
1.4662e+005	1.4662e+005
3.0428e+005	3.0428e+005
2.8624e+005	2.8624e+005
2.0736e+005	2.0736e+005
1.6974e+005	1.6974e+005

Table 5 – comparing covariates means for PSM match (Table 6, 3rd row)

Variable	Treated	All Untreated	Matched Untreated	P-values of Treated versus Matched	P-values of Treated versus All Untreated
R1	0.2439	0.0943	0.2033	0.5550	0.0021
R2	0.0976	0.0821	0.1220	0.6559	0.7293
R3	0.0000	0.1533	0.0000	NaN	0.0073
R4	0.1707	0.1643	0.1829	0.8517	0.9140
R5	0.1951	0.1805	0.1992	0.9520	0.8138
R7	0.0488	0.0873	0.0407	0.8105	0.3944
DSJ2	11738.9080	12686.1231	12154.5148	0.8384	0.5864
DSJ	92.5821	100.6826	94.1589	0.8707	0.3243
DNR2	27.3170	32.0941	28.9991	0.7750	0.6537
DNR	4.3823	4.1390	4.5509	0.7291	0.6945
DLR2	12.1030	10.5939	12.9658	0.7589	0.7065
DLR	2.8501	2.4509	2.9643	0.7406	0.2482
DNP2	304.6311	464.0022	323.0566	0.7826	0.1200
DNP	14.0342	17.4462	14.9657	0.5831	0.0931
DRW2	15.7378	19.0356	16.9801	0.7183	0.5857
DRW	3.3053	3.4062	3.4527	0.6978	0.8176
PTC2	0.3327	0.1111	0.2364	0.3187	0.0011
PTC	0.3662	0.2035	0.3287	0.5524	0.0002
SDA	7.1184	6.4256	7.1832	0.9591	0.5562
REL2	0.4555	0.4045	0.4150	0.7154	0.7358
REL	0.5175	0.4113	0.4805	0.6112	0.1744
RAI2	15.4878	12.6653	14.7119	0.6247	0.0131
RAI	3.7073	3.4335	3.6492	0.7753	0.0774

Table 6 – comparing results (matching uses n=6) (* = drop points with poor matches)

	No Bias Adjustment	No Bias Adjustment	Bias Adjusted	Bias Adjusted
	3-year effect (%)	Annual effect (%)	3-year Effect (%)	Annual effect (%)
All Data	-0.63 (-0.51)	- 0.21	-0.42 (-0.33)	- 0.14
PSM	-0.29 (-0.30)	- 0.10	-0.46 (-0.46)	- 0.15
PSM *	-0.43 (-0.41)	- 0.14	-0.48 (-0.45)	- 0.16
CM	-0.40 (-0.81)	- 0.13	-0.00 (-0.02)	- 0.00